3-17-2020

The role of supplier performance evaluations in mitigating risk: assessing evaluation processes and behaviors

Timothy G. Hawkins

Michael J. Gravier
Bryant University, mgravier@bryant.edu

William A. Muir

Contents lists available at ScienceDirect

Industrial Marketing Management

journal homepage: http://ees.elsevier.com

The role of supplier performance evaluations in mitigating risk: Assessing evaluation processes and behaviors

Timothy G. Hawkins [a], Michael J. Gravier [b,*], William A. Muir [c]

[a] Logistics, Department of Marketing, Logistics, & Operations Management, G. Brint Ryan College of Business, 1155 Union Circle #311396, University of North Texas, Denton, TX 76203, United States of America
[b] Bryant University, 1150 Douglas Pike, Smithfield, RI 02917, United States of America
[c] Naval Postgraduate School, 555 Dyer Rd, Monterey, CA 93943, United States of America

ARTICLE INFO

Keywords
Supplier performance evaluation
Opportunism
Past performance
Agency theory
Organizational behavior
Procurement

ABSTRACT

Given the level of outsourcing, supplier performance evaluation (SPE) is a critical supply chain process. SPEs are used to record supplier performance levels to inform future supplier selections, and thus mitigate the risk of adverse selection. Numerous weaknesses associated with industrial buyers' collection and use of supplier performance information call SPE effectiveness into question. The risk-related factors affecting SPE effectiveness have not been empirically explored, including misuses of the tool. This research identifies the factors affecting SPE risk mitigation effectiveness. It employs a mixed method of qualitative interviews of buyers and suppliers in order to develop a model of SPE risk mitigation effectiveness using structural equations modeling of survey data from a rare sample of 131 performance assessors. Findings implicate the importance of a thoroughly defined scope of work, an accurate SPE, and documented rating justifications. Additionally, dissonance among several performance evaluators and the fear of a supplier's dispute detract from SPE risk mitigation effectiveness. Finally, this research unveils how SPEs are weaponized, pursuing short-term gains and clouding the view of the supplier's performance thereby hindering the long-term, risk-mitigating purpose of SPEs. Two separate forms of opportunism - *threat* and *debt* - are discovered and have differing effects.

1. Introduction

Given the level of outsourcing, supplier performance evaluation (SPE) is a critical supply chain process. SPE is "the process of evaluating, measuring, and monitoring supplier performance and suppliers' business processes and practices for the purposes of reducing costs, mitigating risk, and driving continuous improvement" (Gordon, 2008, p. 4). Measuring supplier quality is critical since the cost of poor quality ranges from 10% to 25% of sales, and the cost of poor *supplier* quality ranges from 25% to 70% of the cost of poor quality (Gordon, 2008). SPE improves buyer–supplier relationships (Prahinski & Benton, 2004; Ulaga, 2003), increases supplier performance (Prahinski & Fan, 2007; Ulaga, 2003), and yields reductions in total supply chain costs (Monczka, Choi, Kim, & McDowell, 2011).

SPE became popular in the 1950s (Wieters & Ostrom, 1979), and now SPE is a ubiquitous (CAPS Research, 2011) and essential best practice in business-to-business sourcing (Gordon, 2008). SPEs inform future supplier selection decisions of the likelihood that a prospective supplier will successfully perform the contract. Assessing and recording performance levels and making that information available to buyers during a future source selection is believed to make suppliers work harder to ensure satisfactory (or better) performance (OFPP, 2000).

SPE schemes are not without challenges. Too often, they are not properly, timely, or accurately completed (Buffa & Ross, 2011; GAO, 2014). Reports often lack sufficient information to support ratings necessary to withstand a challenge, or do not include a rating for all performance areas (OFPP, 2011). Additionally, throughout the rating process, raters are sometimes inclined to inflate ratings in order to avoid conflict with the supplier (GAO, 2009). In fact, avoiding purchasing dissonance in order to alleviate psychological distress appears to be an important component of purchasing manager behavior (Gonzalez-Padron, Hult, & Calantone, 2008; Reilly, Saini, & Skiba, 2018). Underlying data captured in information technology systems are sometimes flawed (Hald & Ellegaard, 2011). Masses of performance data are condensed into more general ratings sacrificing fidelity (Hald & Ellegaard, 2011). Buyers also commonly use multiple evaluators (Buffa & Ross, 2011; Hald & Ellegaard, 2011), which invites different perspectives on supplier performance (Hald & Ellegaard, 2011). Evaluating supplier performance is often more subjective than might be imagined (Gordon, 2008), depending heavily on the devel-

* Corresponding author.
E-mail addresses: timothy.hawkins@unt.edu (T.G. Hawkins); mgravier@bryant.edu (M.J. Gravier); wamuir1@nps.edu (W.A. Muir)

https://doi.org/10.1016/j.indmarman.2020.03.004
Received 5 July 2019; Received in revised form 7 March 2020; Accepted 7 March 2020
Available online xxx
0019-8501/

opment of a common language and set of institutions to regulate and adjudicate interactions (Vargo & Lusch, 2011). Thus, SPEs change throughout the evaluation process through a negotiation process between the buyer and supplier (Hald & Ellegaard, 2011). Notwithstanding, given the importance of SPEs to a supplier's ability to win future business, buyers could opportunistically use SPEs as leverage to extract concessions from suppliers – as they do in other contexts of buyer-supplier negotiations (e.g., changes, contract formation, etc.).

Despite the fallibility of SPE schemes, there are no known studies that quantitatively explore their degree of accuracy or susceptibility to opportunism. Further investigation is needed in order to explore the validity of SPE processes because SPE assessments can affect key outcomes such as future business awards, contract compliance, performance-based payments, supplier reputation, incentive awards, and status achievement. While research has addressed the effect of SPE on supplier performance, the effectiveness of SPEs in assisting buyers' future source selection decisions is questionable (Berrios, 2006; Ulaga, 2003). In other words, we do not know the extent to which SPEs validly build the buyer's confidence in its assessment of the risk of doing business with a particular supplier ex ante.

The purpose of this research is to explain the risk-related effectiveness of SPEs and explore the extent to which the supplier performance information collection and usage achieve the intended goals of mitigating the risk of adverse selection. The following research questions are explored:

1. What factors contribute to the risk-related effectiveness of SPEs?
2. How are SPEs misused?

Answers to these research questions are essential because inaccurate SPEs can harm suppliers' reputations and can bias source selection decisions resulting in adverse selection. If supplier performance information is unreliable, and if buyers and evaluators cannot use the information to discriminate between competitive proposals (Kelman, 2010), the effort of collecting and reporting the supplier performance information is squandered. Notwithstanding, buying organizations often use SPE information to rank suppliers and to determine *preferred supplier* status. The rankings and status are suspect if the underlying SPE ratings are not accurate, resulting in the design of a less-than-optimal supply chain, and increasing the level of psychological dissonance in purchasing professionals (Gonzalez-Padron et al., 2008; Reilly et al., 2018).

The remainder of this paper is organized in the following manner. It begins with a brief background of SPEs and the importance of SPE effectiveness. Then, theories relevant to SPE are summarized, namely agency theory and organizational behavior. To explore the antecedents of SPE risk mitigation effectiveness, this research employs qualitative interviews of buyers and suppliers to develop a conceptual model and testable hypotheses. Next, the study presents the methodologies of quantitative data collection and analysis to test the emerged model. Lastly, discussion, limitations, implications, future research directions, and conclusions are offered.

2. Supplier performance evaluations

A growing body of research addresses four streams of SPE inquiry, namely the functions of SPEs, communications, metrics, and outcomes. First, SPEs serve several functions such as: strategy formulation and clarification, management information, communication with suppliers (e.g., ex ante performance expectations and ex post feedback), communications between departments, decision making and prioritizing (e.g., supplier selection and prioritizing supplier improvement activities), coordination and alignment, motivating suppliers, continuous improvement, and documenting historical performance for use in negotiations (Schmitz & Platts, 2003; Schmitz & Platts, 2004). Importantly, SPEs should also improve suppliers' capabilities, thereby benefit the buyer (Hald & Ellegaard, 2011).

Second, Maestrini, Maccarrone, Caniato, and Luzzini (2018) explore the communication of SPEs from the perspective of signaling theory, concentrating on information shaping, reactions from suppliers, and impacts to buyer-supplier relationships. Hald and Ellegaard (2011) also focus on communication and information exchange, highlighting supplier reactions and how SPE information is negotiated between buyers and suppliers.

A third stream of research concentrates on performance metrics and measurement methods. Simpson, Siguaw, and White (2002) identify the different metrics used in SPE. Buffa and Ross (2011) explore the role of evaluation team diversity on supplier performance using Data Envelopment Analysis. Cormican and Cunningham (2007) developed an evaluation tool considering quality, on-time delivery, and total cost of quality. Other research explores the tenets of relational exchange underpinned by social network theory (Granovetter, 1985). Cousins, Lawson, and Squire (2008) examined how the role of socialization mechanisms in the SPE process affected buyer performance. Giannakis (2007) developed a framework for a SPE method that evaluates supplier relational constructs such as trust, commitment, power, and involvement.

Finally, a body of research focuses on the outcomes of SPEs. Managers have strong perceptions that SPE can lead to several benefits (e.g., Hald & Ellegaard, 2011; Handfield, Cousins, Lawson, & Petersen, 2015; Nair, Jayaram, & Das, 2015). The use of an SPM system improves buyer–supplier relationships (Prahinski & Benton, 2004), improves supplier performance (Schmitz & Platts, 2004), and increases frequency and content of feedback, which also increases supplier performance (Handfield & Nichols Jr, 2004; Nair et al., 2015; Prahinski & Fan, 2007). Additionally, the use of a performance evaluation program increases the strength of the relationship between suppliers' process innovativeness and the buyer's performance benefits (Azadegan, 2011). In one industry benchmark, use of an SPM system garnered 10% greater price savings, 12% better on-time delivery improvement, four times greater quality improvement, and 4% greater improvement in service (Aberdeen Group, 2005). SPM enables continuous improvements that can yield a 3% to 6% reduction in total supply chain costs via continuous improvements (Gordon, 2008).

Industrial marketing research extols the role of monitoring in order to manage the hazards of exchange, specifically, supplier opportunism ex post and adverse selection ex ante (Rindfleisch & Heide, 1997). SPE serves both purposes by measuring supplier performance and by serving as a repository of performance information for future source selections. While SPEs have many post-contractual uses as previously described, we limit the focus of this research to the effectiveness of SPEs in mitigating the risk of adverse selection to coincide with the problems identified and the research questions. Otherwise, attempting to address all of the functions of SPEs would require many more explanatory factors in the model compromising the executability of the study.

SPE risk mitigation effectiveness represents the extent to which SPEs achieve the goal of mitigating the risk of unsuccessful contract performance (i.e., avoiding adverse selection). SPEs mitigate risk by informing future buying teams of a prospective supplier's true past performance levels. While context-specific implementation and theoretical perspectives publish fairly frequently, especially case studies of new supplier evaluation methods (e.g., Hu, Chiu, Yen, & Cheng, 2015; Zeydan, Çolpan, & Çobanoğlu, 2011), empirical validations of benefits are scarce (Maestrini et al., 2018). Furthermore, the mechanisms for capturing and measuring performance levels are not well understood, especially in relation to the effects of opportunism.

Our research, and prior research, alludes to the misuse of SPEs (Hald & Ellegaard, 2011). We define *SPE misuse* as the use of SPEs in improper, unforeseeable, or unintended manners. Business research has a history of highlighting the dangers of misuse, for instance in consumer misuse of products (e.g., Trombetta & Wilson, 1975), in institutional misuse of personal information (e.g., Young & Quan-Haase, 2013), and in employee misuse of information technology resources (e.g., D'Arcy & Devaraj, 2012). SPE misuse inherently involves opportunistic behaviors by actors who find uses for SPEs that are

not proper, foreseen or anticipated, and that ultimately erode SPE efficiency and effectiveness at accurately evaluating supplier behavior.

3. Theoretical foundation and hypothesis development

3.1. Multi-theoretical approach

Inter-organizational behavior is a complex phenomenon such that any single theory falls short of explaining the many interrelationships. A review of supply chain literature revealed 180 distinct theories used to direct inquiry and understand the domain (Defee, Williams, Randall, & Thomas, 2010). SPE's complexity emerges from the interaction of individual human actors nested in organizational structures, and due to the intersection of multiple disciplines such as contract law, supply chain management, and marketing. An SPE system is an organizational adaptive mechanism that suffers the problems of agency arising from the managers both internal and external to the firm in the supplier evaluation process (Schmitz & Platts, 2004). Due to its emergence from both the organizational and agent levels, we apply two theories: organizational behavior and agency theory.

Blending organizational behavior and agency theory is characteristic of the field of organizational economics (OE) (Barney & Ouchi, 1986). OE focuses on the relationships between suppliers and buyers, especially the question of what information is gathered and how it is used (Gibbons & Roberts, 2013). Our research falls in the same vein as other buyer-supplier research connecting the inter-organizational and inter-personal levels (i.e., Ashnai, Henneberg, Naudé, & Francescucci, 2016).

OE applies agency theory to study transactions occurring within firms with a special focus on the problem of the shirking and deceit of managers entrusted to protect the interests of the firm versus the managers of the supplying firms (Donaldson, 1990). As a body of empirical and theoretical work, OE presents a stream of organizational behavior literature explaining why some organizations outperform others based upon their ability to manage both internal and supplier relationships efficiently (Barney & Hesterly, 2006). Incorporating an organizational-level theory shores up agency theory research that suffers "misplaced methodological individualism" (Worsham, Eisner, & Ringquist, 1997, p. 423) where large organizations of many people and sub-organizations are assumed to act as one homogeneous entity.

OE explains the need to monitor and control opportunism as much for internal as external managers by viewing organizational behavior in the context of conducting exchanges of goods and services as a continuously evolving process guided by the agents' (internal and external managers') drive to do so in the most efficient manner. Williamson is arguably OE's most famous proponent, and his Nobel citation recognized the implications of his work going beyond his famous transaction cost analysis to bring together economics, organization theory, and contract law, with transaction costs as his most famous explanatory mechanism for assessing intra- and inter-firm efficiency, and especially the effects of opportunism such as shirking, and deceit (Combs & Ketchen Jr, 1999; Donaldson, 1990). Contractual relations engage authority and incentives not available in open market transactions in order to manage divergent goals, imperfect information, and self-interest (Combs & Ketchen Jr, 1999; Williamson, 1975), and effective SPEs contribute to these goals.

Although contracts and organizations create the environment and the processes for SPEs, organizations are constituted of individuals and it would be an error to attribute every positive or negative outcome to the organizational or environmental context (Luthans & Youssef, 2007). Where organizational behavior examines organizational adaptations, agency theory elucidates situations when individuals make decisions on behalf of others, and like SPEs, agency theory's usual unit of analysis is the contract (Eisenhardt, 1989). The contract demonstrates the added complexity of an agent monitoring an agent—the performance evaluator acts as an agent in the employ of the organization, and the supplier is a contractually obligated agent. Asymmetries in information can lead either or both agents to shirk duties, distort information, and behave opportunistically. For example, when rating a sub-par supplier, rather than rate the supplier as unsatisfactory, the evaluator might inflate the rating to avoid a dispute. Principals can respond by increasing monitoring of agents (although aligning the goals of the agent with the principal is a less costly approach), manifesting as long-term employment and, for suppliers, using outcome-based contracts (Eisenhardt, 1989). Principals also screen potential supplier agents ex ante to mitigate adverse selection.

Because these are organizational and agent-level phenomena, our hypothesis development describes the context of the circumstances of our sample setting of government purchasing. The circumstances of government contracting may vary in detail, but we observe similar circumstances in the for-profit setting (Beausoleil, 2010).

3.2. Rating dissonance

SPE is complex; there may be multiple internal stakeholders and organizations, multiple supplier personnel, multiple performance criteria, and multiple performance evaluators involved (Hald & Ellegaard, 2011; Maestrini et al., 2018; Nair et al., 2015; Palmatier, 2008; Shapiro, 2005; Wieters & Ostrom, 1979), especially with large, complex contracts and where performance occurs in more than one location. Evidence of the benefit of reducing organizational dissonance can be found in supply chain integration research. Firm internal integration reinforces external supply chain integration which, in turn, improves supply chain performance (Frohlich & Westbrook, 2001; Huo, 2012; Leuschner, Rogers, & Charvet, 2013).

Dissonance can arise from psychological factors at the agent-level. Affective trait diversity on the source selection team has recently been found to diminish performance of sourcing teams, and sourcing team cohesion fully mediated this effect, demonstrating the importance of reducing dissonance and reconciling diverse sourcing team members (Kaufmann & Wagner, 2017). Dissonance can also be seen in the negative effect of geographic dispersion on buyer-supplier relationship performance that increases with the intensity of information sharing (Lorentz, Töyli, Solakivi, Hälinen, & Ojala, 2012), findings mirrored in international trade (Disdier & Head, 2008).

Rating dissonance represents one component of information integration in supply chains (Frohlich & Westbrook, 2001). Organizational behaviorists have researched dissonance among multiple raters in the context of the employee performance appraisal system (e.g., 360-degree evaluations in which superiors, subordinates, and peers evaluate the ratee). Using multiple raters can offer more unique, valuable information about the employee's performance and may mitigate evaluation bias (Brown, Inceoglu, & Lin, 2017; Levy, Cawley, & Foti, 1998; van der Heijden & Nijhof, 2004). More fairness comes at the cost of variance in ratings attributable to individual differences in raters (Brown et al., 2017; Mount, Judge, Scullen, Sytsma, & Hezlett, 1998). Notwithstanding, a rich research stream touts the benefits of functional conflict - specifically task conflict (Jehn & Chatman, 2000) - which facilitates challenges to ideas and open dialogue. Managed correctly, such disagreements improve decision-making and increase performance (Jehn & Chatman, 2000).

In the context of SPEs, friction results from the organization concatenating insights from different evaluators, different instances of supplier performance, and different interpretations of the meaning of performance criteria and rating definitions (Dowst, 1972; Korschun, 2015; Levy et al., 1998). More raters mean more complexity to input ratings and assessments, and more opportunity for suppliers to successfully rebut ratings and to offset relatively minor failures with successes, garnering an overall acceptable rating. An increased ability to escape a rating unscathed diminishes motivation to improve performance, and reduces negative performance information visible in future source selections.

With both beneficial and detrimental effects, it is likely that dissonance affects SPE efficacy; however, it is not feasible to predict exactly how. Therefore:

$H1_a$There will be a negative relationship between rating dissonance and SPE risk mitigation effectiveness.

$H1_b$There will be a positive relationship between rating dissonance and SPE risk mitigation effectiveness.

3.3. Rating justification and perceived accuracy

In performance appraisal, suppliers will more willingly accept negative feedback based on detailed, thought-out communication and clearly identified initiatives and improvements (Maestrini et al., 2018). Justifications include the supporting details and facts surrounding the evaluated performance, applying both mediated and non-mediated (referent, expert) bases of power to create an ethical buyer-supplier climate (Schleper, Blome, & Wuttke, 2017). Fact-based substantiation and commitment to an ethical climate give credence to the rating and give confidence to the future source selection team that the SPE information is providing true insights into the prospective supplier's past performance. *Rating justification* describes when the source selection team is armed with the facts to support their ratings and to defend them to senior leaders or even against unsuccessful offerors in a legal challenge to a contract award decision (e.g., a bid protest in government procurement).

No matter how well justified, *perceived accuracy* of SPEs depends on several factors. Buying organizations that cannot muster the evidence to justify a rating, or opt not to bother, diminish accuracy. Inaccuracies also creep into SPEs in ways that demonstrate striking parallels to employee performance appraisal. Affective constraints limit the amount of agreement between a supervisor's rating and a ratees' self-evaluation due to the tendency of individuals to confirm a role in successful performance while attributing failure to external factors (Campbell & Lee, 1988; Deb, Li, & Mukherjee, 2016). In the context of organizational buying, failures of the procurement program could be unreasonably attributed to a supplier's performance.

At other times, the employee or supervisor willfully gives an inaccurate appraisal, perhaps in order to preserve the effectiveness of an interdependent work group (Campbell & Lee, 1988; Korschun, 2015). Academic literature confirms that a halo effect occurs in SPEs and could partially explain inflated ratings (Kelman, 2010), and the same concern has been identified in employee performance appraisals that demonstrate similar traits of multiple raters and inflated ratings (Erez, Schilpzand, Leavitt, Woolum, & Judge, 2015; Thomas & Bretz, 1994). In another situation that occurs in the supplier relationship context, supplier self-evaluations may become biased if a supplier seeks to preserve its reputation; deliberate dishonesty is more likely to occur when self-appraisals are used for scarce resource allocation decisions (Shrauger & Osberg, 1981; Steinle, Schiele, & Ernst, 2014), such as contract awards.

Distressed suppliers may refute any negative information, challenging the rating and justification, increasing workload for the buying organization to resolve disagreements. If the supplier's performance did not meet requirements, rather than rate the supplier as unsatisfactory, the evaluator might inflate the rating to avoid a dispute—conflict that would unveil the evaluator's negligence.

In sum, SPE systems with deficient *ratings justification* and *perceived accuracy* disincentivize suppliers and fail to reduce the risk of adverse selection by a future buyer (Maestrini et al., 2018). Perceiving a deficient system, evaluators will fail to engage in detailed, factual rating justification that will be accepted by the supplier and, if rebutted, internally by a reviewing official. Thus, it is posited that:

H2There is a positive relationship between perceived accuracy and rating justification.

H3There will be a positive relationship between rating justification and SPE risk mitigation effectiveness.

3.4. Supplier disputes

Perceived unfairness has been labeled as inter-firm "relationship poison", and while contracts suppress problems of opportunism they exacerbate the sense of unfairness due to attribution at the agent-level of negative motivations to the evaluator (Samaha, Palmatier, & Dant, 2011). This mirrors organizational behavior research about how a ratee will respond to feedback in a way that corresponds to his or her acceptance of the feedback: "employees are unlikely to accept, desire to respond, or intend to respond to feedback based on information derived from an invalid or inaccurate appraisal" (Kinicki, Prussia, McKee-Ryan, & Wu, 2004, p. 1067). The emergence of due process in employee performance evaluation is testament to the importance of a fair hearing (i.e., an appeal process) and that judgment be based on evidence (i.e., be accurate) (Folger, Konovsky, & Cropanzano, 1992).

In government contracting, ratings are entered into an information system by *assessing officials* – those appointed to oversee supplier performance (e.g., quality inspectors). The frequency of evaluation varies by contract, but typically, an evaluation occurs within 120 days of the completion of performance (e.g., delivery of goods or completion of a term for services). The information system transmits the evaluation to the supplier. Suppliers are afforded an opportunity to respond to the evaluations. Disagreements are decided by a *reviewing official* who serves at least one level above the assessing official. Final reports are again shared with the supplier.

Disputed ratings generate efforts to justify the original ratings (Beausoleil, 2010). Rebuttals often occur in SPEs in for-profit organizations as well (Hald & Ellegaard, 2011); a supplier may leverage a manager-level relationship with the buyer to communicate disagreement with SPEs. Any rebuttals require that the buying organization develop internal explanations, expend effort to muster facts, and deal with negative attention if the original SPE was not accurate or properly justified. *Fear of a supplier dispute* can result when attempts among multiple raters to thwart a supplier rebuttal generate internal conflict, which may motivate some evaluators to inflate ratings—and diminish ratings accuracy—to avoid a dispute, while others may take a legalistic, strict approach. Given the above logic, it is hypothesized that:

H4The lower the perceived accuracy, the greater the fear of supplier dispute.

H5There will be a positive relationship between fear of supplier dispute and rating dissonance.

3.5. Sufficiency of requirement definition

Buyer expectations of suppliers must be communicated to suppliers. This often occurs via a requirements document such as a statement of work, a specification, or drawings. Purchasers of goods and services should clearly define their requirements (i.e., expectations) to achieve procurement objectives (Van der Valk & Rozemeijer, 2009). This is among the most difficult tasks in the sourcing process (Van der Valk & Rozemeijer, 2009). Reports highlight several instances of decreased contract outcomes due to inadequately defined requirements (GAO, 2002; GAO, 2007; GAO, 2009). These persistent problems illuminate the problems of agency. Defying conventional wisdom, buyer teams fail to apply the requisite due diligence to develop and provide complete and accurate specifications of needs. Without a complete understanding of the buyer's requirement, a supplier may not perform work that the buyer expects to receive and may not meet the buyer's expectations in terms of function, performance, and quality (Hawkins & Muir, 2014). Logically, if the work is not well defined, evaluators will struggle to accurately evaluate the supplier's efforts.

Therefore, is it posited that:

H6There will be a positive relationship between the sufficiency of the requirement definition and perceived accuracy.

3.6. Opportunism attitudes

Central to agency theory is the assumption of opportunism, defined as self-interest seeking with guile (Williamson, 1975). Opportunism constitutes a litany: stealing, cheating, breach of contract, dishonesty, distorting data, obfuscating issues, and misrepresentation (Anderson, 1988; John, 1984; Wathne & Heide, 2000; Williamson, 1981, 1987, 1993). It is so commonplace that textbooks describe opportunistic negotiation techniques such as phantom offers, escalation, the switch, silence (Monczka, Trent, & Handfield, 2002), artificial legal leverage, the missing person, stalling (Cavinato & Kauffman, 2000), and bluffing.

Where the positive aspects of trust and long-term relational orientation have enjoyed research attention, little research has assessed the opportunistic manner in which buyers can take advantage of their important role in the buyer-supplier setting (Schoenherr et al., 2012), and even fewer works have assessed opportunism in the context of SPE misuse. Hald and Ellegaard (2011) studied SPE's in their design, implementation, and use phases, and identified 13 factors that shape supplier performance information. Their work found potential misuse by the suppliers at the implementation phase, and by both suppliers and buyers at the use phase.

The misuses resulted from amplifying effects at the implementation stage, and from dampening and directing effects at the use stage. Amplification resulted when buyers added or dosed information to provoke a behavioral change—behavior that left suppliers feeling treated unfairly and demotivated. Dampening is a withdrawal behavior involving reduction of a past amplification in order to restore goodwill, but it may drive supplier doubts in the SPE, thereby reducing their effectiveness. Directing defines the effect of re-routing SPE information to influence who will receive it, or blocking or hindering information from reaching its intended audience.

Dependence is the predominant factor affecting opportunism (Hawkins, Wittmann, & Beyerlein, 2008; Tangpong, Li, & Hung, 2016; Wang & Yang, 2013). Empirical research finds variable impacts on trust from different forms of buyer opportunism. Hill, Eckerd, Wilson, and Greer (2009) distinguished subtle and deceitful unethical behaviors, finding that even if the supplier benefits from buyer unethical behavior that trust erodes—a supplier dependent on a buyer untrustworthy in one dimension would understandably exhibit less trust in other dimensions of the relationship, such as SPEs.

The extremely high switching costs of engaging in government contracting also create dependence of buyers on suppliers. The oft-used sole source contract increases buyer dependence (and supplier power). In order to counter dependence, buyers may weaponize the SPE as leverage to reap concessions from suppliers. The practice of a buyer threatening to downgrade a supplier's performance rating unless a particular action (or inaction) is conceded (Schmitz & Platts, 2004) constitutes deceitful psychological contract violation (Hill et al., 2009) hereafter referred to as *opportunism-threat*. This opportunism manifestation is misusing SPEs for coercion by withholding rewards (of a deserved more favorable SPE) versus by levying punishment (Molm, 1997; Saini, 2010; Tangpong et al., 2016).

Past research demonstrates that where breaches of relational expectations can be somewhat overlooked when attributed to environmental circumstances beyond buyer control such as a disruption, when attributed to reneging such behavior particularly damages trust (Eckerd, Hill, Boyer, Donohue, & Ward, 2013), and reduced trust erodes supplier confidence in the buyer's performance evaluation process. The subjective nature of many performance evaluation schemes (Gordon, 2008) provide buyers ample latitude to conclude performance as acceptable or not.

On the other hand, a less-than-desirable performance rating may be warranted, but a buyer may be willing to compromise the long-term value of the SPE (i.e., the ability to inform a future source selection) by inflating the rating in exchange for near-term concessions on the current contract (Husser, Gautier, André, & Lespinet-Najib, 2014; Tangpong et al., 2016). Where ratings are subtly bargained for concessions, the accuracy of SPEs could be questioned. Therefore, it is posited that:

H7Opportunism Attitude-Threat will be negatively related to perceived accuracy.

Opportunism in SPEs could take another form. In a situation that supplier performance warrants a 'poor' SPE rating, fear of a supplier dispute to the rating would be high because suppliers often will attempt to challenge any information deleterious to its reputation – whether the basis of the challenge is true or not (Maestrini et al., 2018). To avoid confrontation, the buyer can award an obviously more positive SPE than deserved. This 'gift' is an SPE misuse that could create a subtle debt and activate the norm of reciprocity (Greenberg, 1980; Tangpong et al., 2016). The long time horizon and mutual buyer-seller dependence means that the buyer can expect a return favor in the future, effectively supplanting the written contract with a psychological one forged by promises between individual agents (Kingshott, 2006). This can dilute accountability between the buyer and supplier, and re-purposes trust from benefiting organizations in order to benefit individual agents (Husser et al., 2014; Saini, 2010).

We refer to this as *opportunism-debt*. The theory of indebtedness (Gouldner, 1960) suggests that debts can create negative emotions such as discomfort, uneasiness, and avoidance mechanisms. Debts restrain freedom potentially leading to reactance (Pelser et al., 2015). This sort of subtle unethical behavior has been found to increase the sense of psychological contract violation in empirical supply chain research (Hill et al., 2009), and such behavior negatively impacts perceptions of fairness (Eckerd et al., 2013). Even in the context of a trusting and transparent relationship without intentional opportunistic intentions, buyers imposing an obligation that exceeds supplier's willingness or capabilities to reciprocate is a misuse of an SPE that erodes trust (Day, Fawcett, Fawcett, & Magnan, 2013); negative attributions of buyer motivations would erode trust even further (Eckerd et al., 2013; Hill et al., 2009).

Suppliers may attempt to break free from the indebtedness by challenging the premise that the SPE is overly-favorable. The buyer may fear that the supplier, in its rebuttal – which becomes an official record – will explicitly call out the buyer's opportunistic attempt to gain power via the tactic of indebtedness. A supplier can share its experience with other clients and suppliers, potentially tarnishing the reputation of the offending buyer. A rebuttal creates more effort to reconcile the conflict, and unveiled favors with strings attached may become embarrassing if revealed. Therefore, it is posited that:

H8Opportunism Attitude-Debt will be positively related to fear of supplier dispute.

The conceptual model (Fig. 1) displays this set of hypotheses to explain SPE risk mitigation effectiveness. The model integrates organizational theory and agency theory to offer explanations for the mechanisms leading to the risk-related effectiveness of SPEs.

4. Methodology

This research employs a mixed method. First, interviews of buyers and suppliers serve to validate the conceptual model of hypotheses suggested by practice and supported by theory (i.e., research question 1). Next, the interviews were used to explore whether and how SPEs are used opportunistically. Following the qualitative model building phase, a follow-on study was designed to collect quantitative data of

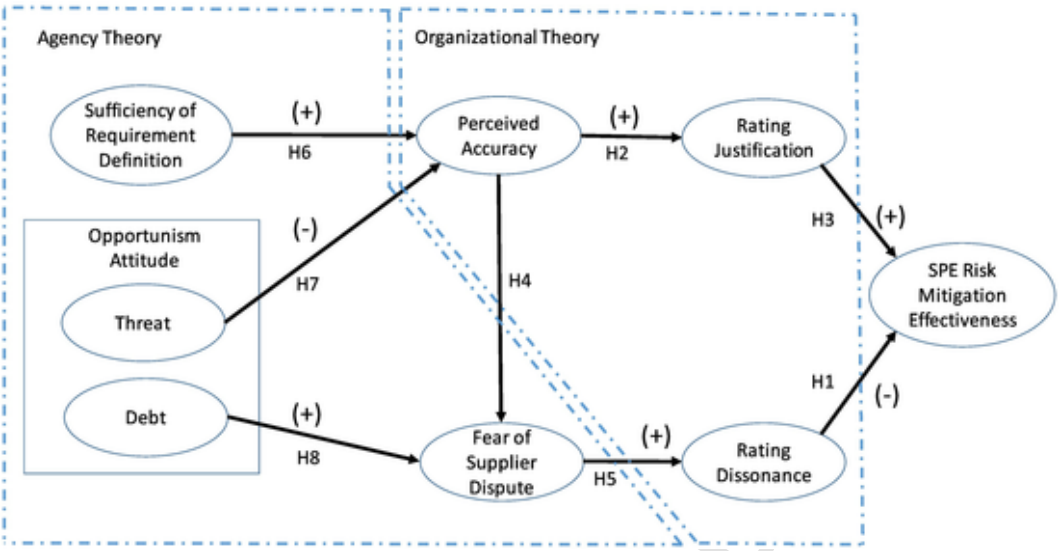

Fig. 1. Model of SPE risk mitigation effectiveness.

SPEs from a rare sample of performance evaluators in order to test the hypotheses.

The context selected for the study was contract performance between U.S. federal government agencies and their suppliers. A government context was suitable due to its massive scope (i.e., dollars, industries, and geographies), rigor, established fairness, and standardized SPE procedures. The U.S. government has a uniform policy of SPE data collection and use, and uses standard ratings. SPE data is considered proprietary or is otherwise sensitive; many for-profit firms have policies prohibiting the release of the information. Taken together, the government context offered opportunity to explore large, complex contracts while controlling for bias and systematic sources of variance.

4.1. Qualitative

Interviews were conducted separately with performance evaluators and suppliers. The interview protocols (Tables 2 and 4) were developed based on a review of archival SPEs, the literature surrounding SPE, and discussions with academic experts and practitioners. Eight interviews of performance evaluators were conducted. Each interview was recorded and then transcribed, lasting between 38 and 67 min (mean of 51 min). Transcripts averaging 18 pages and 7394 words in length were then sent to informants for an accuracy check, enhancing construct validity (Flint, Woodruff, & Gardial, 2002; Yin, 2009). The sample of performance evaluators (Table 1) was drawn from the researcher's personal contacts within one military service. Employees who routinely evaluate contractor performance participated. Experience in evaluating contractor performance ranged from two to 28 years, and there was a similar wide range of the number of SPEs experienced (1–50).

Eight interviews were conducted with suppliers. The interviews lasted between 32 and 65 min, and all were recorded and then transcribed except for two at the request of the informants. Transcripts averaged 13.5 pages in length. One interview occurred in-person, five occurred via telephone, and two informants provided only written testimony. The sample of supplier informants (Table 2) was identified from awarded contracts exceeding $150 thousand and from contacts made at a trade association annual conference. Input from representatives of federal contractors who had been directly involved in the SPE process was sought. The perspectives of large and small businesses representing multiple industries were obtained. Experience in managing customer evaluations ranged from three to 34 years, and there was a similar wide range of the number of SPEs experienced (12–50).

The analysis process began by identifying constructs, defining those constructs, and then positing relationships between them (Van Ecke, Skouma, Freund, Goeskjaer, & Ooms, 2006). Each interview was examined to identify themes and then tested to determine whether these themes remained consistent in subsequent interviews or in reexaminations of previous interviews. Tables 3 and 4 show relevant summaries of testimonies and informant quotes, then identify associated constructs from the conceptual model. The relationships among constructs in the conceptual model were supported by the interview data.

4.2. Quantitative

The quantitative methodology entailed measuring constructs via survey. The unit of analysis was a buying organization's SPE. Existing scales with established reliability and validity were used where possible (Appendix A). Measures were created for constructs with no exist-

Table 1
Informant demographics – performance evaluators.

Informant	Civilian/military	Industry	Experience (Years)	Role	Past performance experience (number of evaluations)
1	Civilian	Aerospace	28	Contracting Officer	50+
2	Military	Aerospace	7	Program Manager	10
3	Civilian	IT	4	Program Manager	11
4	Civilian	IT	10	Program Manager	7
5	Military	IT	10	Program Manager	5
6	Military	IT	9	Program Manager	15
7	Military	IT	2	Program Manager	1
8	Military	IT	18	Program Manager	10

Table 2
Informant demographics – suppliers.

Informant	Business size	Industry	Experience managing customer evaluations (years)	Duty title	Supplier performance experience (number of evaluations)
1	L	Aerospace	34	Systems Engineer	Multiple
2	L	Aerospace	7	Program Manager	50
3	L	Aerospace	14	Program Manager	Multiple
4	S	Information Technology	3	COO	
5	L	Munitions	4	VP, Business Development	12
6	L	Aerospace	30	VP, Business Strategy	50
7	L	Aerospace	Multiple	Contracts Director	24
8	S	Shipbuilding	30	President	30 +

ing scales based on the literature and on comments from performance evaluators and suppliers. New scales were used to measure SPE risk mitigation effectiveness, rating dissonance, rating justification, opportunism attitude-threat, opportunism-debt, and fear of supplier dispute.

4.2.1. Perceived accuracy

Perceived accuracy was measured by expounding on a scale developed by Kinicki et al. (2004) that assessed the accuracy of employee performance appraisals. Key aspects of accuracy included being consistent and being factual. Informants reported the importance of factual evaluations since a supplier's future business could depend on the SPEs. They also confirmed the *halo effect* (Erez et al., 2015); supplier evaluations are commonly inaccurate due to inflated ratings. Informants also expressed concern about inconsistent SPEs due to the human factor and the inherent subjectivity of ratings. Therefore, two items each were used to measure these two aspects.

4.2.2. Sufficiency of requirement definition

An existing, four-item scale was used to measure the sufficiency of requirement definition (Hawkins, Berkowitz, Muir, & Gravier, 2015). Requirements (i.e., needs) are commonly communicated in various documents such as contracts, purchase orders, specifications, drawings, and statements of work; thus, several scale items prompted respondents to consider these documents to indicate how well the requirements were defined. Often, these documents are not perfect (Lam, Chin, & Pun, 2007). Ambiguities in the wording can lead to uncertainty in meaning. Thus, one item was included to gauge the presence of ambiguities.

4.2.3. Rating dissonance

This construct gauges the degree of disagreement among multiple performance evaluators of a supplier's performance level. As evidenced in the interview testimonies, consternation can envelop ratings and narrative explanations of the ratings. Thus, three scale items addressed these two potential points of disagreement. Consistent with prior operationalization of disagreement (Lovelace, Shapiro, & Weingart, 2001), scale items asked respondents to indicate the amount of disagreement and lack of consensus.

4.2.4. Rating justification

Behavioral decision-making research emphasizes the important role of reasoning in choices (Kuo & Nakhata, 2016). As such, rating justification addresses the extent to which the evaluation was explained and warranted. Justification taps an important aspect of SPE, acceptance of feedback (Ilgen, Fisher, & Taylor, 1979). Buyers (i.e., raters) want suppliers (ratees) to accept the evaluation in order to avoid conflict and motivate them to appropriately act on the feedback. In the context of SPEs – and unique to a buyer-supplier context – the buying organization also needs future buyers to accept the feedback. Acceptance by future buyers enables trust in the information, and thus, buyers can confidently rely on (i.e., act upon) evaluations that are explained and supported with evidence. Decision justification has been operationalized as easy to defend (i.e., evidence) and logical (i.e., explained and rationalized) (Kuo & Nakhata, 2016). These key aspects of justification emerged in the interviews; thus, two scale items were included to assess the extent of explanation of the ratings, and one item addressed the extent of evidence to support the SPE.

4.2.5. Opportunism

Interviews with performance evaluators and suppliers revealed opportunistic tendencies among evaluators. The willingness to use the threat of a low SPE to evoke a desired response from a supplier (referred to as *opportunism attitude – threat*) was measured by two items. One item assessed the acceptability of threatening a supplier using the SPE rating, while the other item assessed the acceptability of using the SPE as leverage. The willingness to award an overly-favorable SPE with the expectation that the favor will be repaid (referred to as opportunism attitude – debt) was measured by two items. The first item gauged the expectation of reciprocating. The second item assessed the evaluator's belief that leverage can be gained through gifting an inflated SPE.

4.2.6. Fear of supplier dispute

This construct emerged from the interviews as a consideration affecting the content of SPEs. Scale items assessed the likelihood of the supplier disputing its SPE, the buyer's degree of concern over such a rebuttal, and the ensuing effort required to resolve the dispute.

4.2.7. SPE risk mitigation effectiveness

Since the purpose of SPEs is to mitigate the risk of adverse selection, gauging the ability of the SPE to assess the risk of unsuccessful supplier performance ex ante is essential. The developed scale included two items to assess such risk. Inherent in the definition of risk is uncertainty (Bettman, 1973) – an aspect that emerged from the interviews. Therefore, one scale item was used to elicit respondent's consideration of the uncertainty of supplier performance. Given the purpose of the SPE to reduce information asymmetry between buyers and prospective suppliers during a future source selection, we added a scale item that gauges the buyer's confidence level in its assessment of potential supplier risk.

As a pre-test, several academicians and contracting practitioners reviewed the survey instrument. Feedback was solicited regarding whether the survey items: (1) captured the domain of the construct (content validity), (2) were unambiguous, (3) were simple to understand, and (4) were consistently interpretable (Dillman, 2000). The experts were asked whether the model was sufficiently comprehensive, and it was modified based upon expert feedback.

In an effort to ensure construct reliability and validity prior to full-scale survey deployment (Churchill Jr., 1979), the survey instrument was pilot-tested via online survey to 265 performance evaluators, with 41 responses received. Internal consistency reliability for each latent construct was assessed using Cronbach's alpha with all constructs show-

Table 3
Structured interview questionnaire – performance evaluators.

Question, Response Summaries, and Quotes	Construct
Are past performance reports useful? How so, or why not? • "I think it could be effective at mitigating a risk if the requirements that you are looking at match up with the [inaudible] past performance evaluations that you are comparing them to." • "I know that it is going to be watered down kind of like the [enlisted performance report/officer performance report] because there is so much pressure that the contractor puts back on the government for wording intricacies. Overall, I think I would have to question the overall overarching fairness of the process just because just like the [enlisted performance report/officer performance report] system, particularly the [officer performance report] system you question how much reality you are getting out of this if you are not seeing all of these support that goes behind the ratings. That is why I would have to say overall I would question it."	SPE Effectiveness Fear of Supplier Dispute Rating Justification
In the cases of multiple evaluators on a single contract action, do past performance evaluations/ratings deviate among evaluators, and, if so, why? • "Sometimes there was some real consternation, and sometimes they actually went outside the program team and went up to higher management to get it resolved." • The informants offered a variety of explanations for differences in assessments. Three informants mentioned different expectations of contractor performance and poor requirements definition as culprits. Two informants attributed incongruent past performance evaluations to insufficient monitoring of the contractor. Two informants mentioned that the different government performance evaluators had different experiences, suggesting that individual differences may exist. Two informants mentioned different locations of the contracting officer's representative, indicating that performance may differ at different physical sites. Two informants also mentioned that work overload precludes performance evaluators from fulfilling their duties to evaluate and document contractor performance.	Rating Dissonance Sufficiency of Requirement Definition Accuracy
To what extent do past performance evaluations/ratings captured in federal databases influence source selection decisions? • One informant reported no influence. Three informants reported little influence. One informant reported some influence, and one informant reported great influence.	SPE Effectiveness
Why do past performance evaluations/ratings lack sufficient justification/supporting information? • "I was working on another project completely different from this and couldn't even spell [SPE]. I mean I didn't really know what it was and all of a sudden I was made the program manager for a certain—for a program." "So my first one was—and I don't even remember what the ratings were—I really don't, but I know that first one, that was probably—I am not going to say it was wrong, but I am going to say it was—I couldn't have backed up some of the stuff that was in there because I wasn't working with the contractor." • "And there is a wide variety within the system, in my experience. So you get—and you find that out by calling back to the PMs that you can get ahold of, if they are still there. The older the [SPEs] are, obviously it is harder to find the people, and you clarify the information you are reading from a past performance perspective."	Accuracy Rating Justification Rating Justification
Why are past performance evaluations sometimes inaccurate? • "That is very hard to get an under satisfactory from what I have seen." • "Many—in my opinion, many of the ratings for a long time could have been a lot lower if government had its act together and adequately supported and communicated with the contractor." • "Some services tend to not put much negative information in there in my experience. At least the ones I have read. Some of them are written more like a performance report where it's bad to say anything negative. I think that—if that is the approach that people take, then you would take then the system has little value." • "There were other things that were like, well, they didn't perform as well as we wanted them to, but we couldn't ding them on it because nowhere in the contract did it specifically say this is your standard and this is where you have to meet it or exceed it." • "So we work hard in this division to have the evidence within the [SPE] so it doesn't get disputed down the road if we run into issues." • "Actually it didn't even get disputed, and we had a couple of areas where we had a few markdowns and we had the data, and that is the important thing in writing is the data to back it up."	Accuracy Accuracy Accuracy Sufficiency of Requirement Definition Rating Justification Fear of Supplier Dispute Rating Justification

ing reliabilities greater than 0.7 for established scales and greater than 0.6 for new scales (Hair, Black, & Babin, 2010).

An online survey was then presented to 2247 additional performance evaluators. Out of 148 responses received, 58 were incomplete resulting in 90 usable responses. Since no changes were made to the survey from the pilot study, the records from the pilot study were added to the data set[1] resulting in a combined sample of 131 respondents out of 2512 and an overall response rate of 5.2%. This response rate, while low, is comparable to that of other supply chain research that combined samples (Gimenez & Sierra, 2013) and other studies of public procurement, a population that is difficult to access (Saastamoinen, Reijonen, & Tammi, 2017; Tammi, Saastamoinen, & Reijonin, 2014). The response rate is also similar to previous studies in marketing research (e.g. Lages, Silva, & Styles, 2009; Morgan & Hunt, 1994; Prior, 2012; Prior, 2016). Due to the sensitivity of SPEs, the survey data is difficult to obtain and, therefore, extraordinarily rare. No other study that we are aware of exists that captures quantitative data of supplier evaluations of multiple suppliers from multiple buy-side evaluators of post-contract-award performance.

4.3. Demographics

Demographics characterizing the respondents and the contracts for which they responded are found in Table 5. The average dollar value of the contracts was $164.7 million (std. dev. $971.8 M; range: $62 K-$10B). The respondents' average years of experience assessing contractor performance was 14.75 (std. dev. 9.5). The sample was respectably educated. Performance evaluators represented a variety of job functions. Respondent ages were evenly distributed across ten-year groups. Most respondents were male (72%), which is somewhat skewed compared to total U.S. government employment (57%) (Office of Personnel Management, 2014). The sample was influenced by services versus construction and goods, but a wide variety of goods and services were represented (36 different product-service codes/federal supply classes). Most contracts were competed, and large and small businesses are evenly represented. All major types of contracts were represented.

[1] We include a control for survey group within our subsequent analysis.

Table 4
Structure interview questionnaire – suppliers.

Question, response summaries, and quotes	Construct
Why are SPEs often inaccurate?	
• "The natural inclination is that even though it to be over the 12 month period, they think about what has happened to them recently and they tend to think about the bad things more than the good things."	Accuracy
• "Sometimes the report reflected some recent event rather than the entire period."	Accuracy
• "Inconsistency given by the human judgment factor. There is too much subjectivity." "It makes reports unreliable."	Accuracy
• "In our experience [the agency] does not follow and blatantly violates published guidelines for filling these out and uses subjectivity to cover up for its own mistakes during project execution."	Accuracy
• "When a quarterly review with the customer comes back four quarters in a row with an exceptional /very good write up its very hard to accept a satisfactory at the end of the performance period."	Rating Justification
• "With the [SPE], I know there's supposed to be some guidelines on, you know, what's acceptable or marginal, or whatever the guidelines are for performance, but it seems to be not really strictly enforced as far as like there's more motion on the ratings of the [SPE] with not a lot of justification."	Rating Justification
• "There are some shortfalls in it, it is definitely subjective. In other words, most especially between different customers - meaning that we have contracts with various contracting agencies across the government. Some are a lot harder raters than others."	Accuracy
• "We have had reports over a year late."	Accuracy
• "It is more—it is easier to commonly hear them, like 'Oh, [SPE] again.' Again, you know, just with appraisals, 'Oh, appraisal time again.' Yeah, they don't look forward to it and so I think for them it is not a priority, it is not a means of measuring."	Accuracy
• "I have got one contract for sure that there is a monthly scorecard. It is the—the government gives us a monthly look at how we are doing and it basically very much you can tie it right to the [SPE]. There is no fuzz on how things are going throughout the year. So that is great."	Rating
• "It is difficult to meet an 'unknown' requirement."	Justification
• "Sometimes it is [in the contract], but they've gotten rid of our incentive fee, so when we don't have an incentive fee threshold and you don't write it into a contract or a performance work statement, we're kind of shooting in the dark." "Just tell us what it takes—what you expect. If we want to get an exceptional, what do you expect us to provide to you? What performance?"	Rating Justification
• "they're trying to hold us accountable for things that aren't even in the contract." "What they wanted and what they bought were two different things. And so I'm getting dinged on things that again I shouldn't be dinged on. It was not in my contract to provide that level of service for talent."	Accuracy
• "There are a few cases where it just seemed like somebody had an agenda. Those never go over well." Another informant corroborated the existence of an agenda stating: "[The evaluator] distorted the evaluation to suit his/her own agenda."	Accuracy
• "The [SPE] process from the Contractor side can be brutal. The Government personnel can abuse the system and mete out punishment with little to no recourse. The Contractor always appears to be in the wrong as they are replying to accusations."	Accuracy
• "The justification for ratings received have been weak to minimal on the Government's part."	Rating Justification
• When asked whether the informant suspected that the government ever uses the SPE rating/evaluation as leverage, one informant answered: "Yes, especially if they don't get along with the contractor's managers." Another informant stated: "Absolutely, we have a client who we are helping now, because the government client is using this to reduce the requests for equitable adjustment." A third informant commented: "that the [SPE] is used to change our position when we negotiate issues and when tough positions are brought to the forefront. 'Well, you know, you only got this on your CPARS, therefore, you need to work harder so you should give us the—.' So you are told to negotiate." "I think it's used as leverage every time we go to negotiate." "It's implied. It's just hideous. It's under the surface." Another commented: "Yeah, I'd say leverage, because I guess I don't know exactly what gets them, but they definitely use them as an opportu-	Sufficiency of Requirement Definition Sufficiency of Requirement Definition Sufficiency of Requirement Definition Opportunism-Threat Opportunism Rating Justification Opportunism-Threat

Table 5
Sample demographics.

Respondent					
Highest education attained			Career field		
Degree Type		Frequency	Group		Frequency
High School		12	Quality Assurance		3
Associates		8	Program Management		50
Bachelors		31	Contracting		18
Masters		74	Engineering		26
Doctorate		2	Logistics		12
			Other		19
Sex					
Type	Frequency	Percentage			
Male	91	71.7			
Female	36	28.3			
Transaction/contract					
Purchase type			**Competition**		
Type	Frequency		Type	Frequency	Percentage
Services	93		Competed	90	70.3
Construction	4		Sole Source	38	29.7
Supplies/Commodities/Spares	17				
Weapon System	1				
Other	13				
Business size			**Type of contract**		
Type	Frequency	Percentage	Type	Frequency	
Small Business	63	49.2	Firm-Fixed Price	77	
Large Business	65	50.8	Cost Reimbursement	38	
			Time and Materials	3	
			Labor-Hour	1	
			Hybrid	11	
			Other	1	

4.4. Assessing non-response bias

Non-response bias was evaluated by comparing responses from early and late respondents (Armstrong & Overton, 1977). A chi-square test showed no difference across a key demographic, gender. Independent samples *t*-tests explored any differences in constructs measured by continuous measures. These results suggest that the sample was not affected by a non-response bias. Socially-desirable response bias is "the tendency to give answers that make the respondent look good" (Paulhus, 1991, p. 17). This natural tendency may obfuscate the truth; thus, SDR can seriously jeopardize the validity of survey research (Nunnally, 1978; Randall & Fernandez, 1991). Since this research entailed a sensitive, ethical component (i.e., opportunism), anonymity was assured to respondents as technique to reduce the respondent's motivation to respond in a socially acceptable way.

4.5. Assessing common method variance

We assess and account for the presence of common method variance (CMV) using the correlational marker variable approach described by Malhotra, Kim, and Patil (2006). CMV refers to systematic error variance resulting from the use of a single (common) method, potentially inflating correlations between variables and obscuring the true magnitude and, possibly, significance of statistical relationships. CMV is of particular concern in cross-sectional survey research, where data on multiple variables are collected from a single source. To partial out this variance, and thus to guard against CMV-related biases in our model-produced coefficient estimates, we utilize the second-smallest positive correlation ($\rho < 0.01$) among manifest variables to correct observed correlations between all variables in our study. We then use the resulting matrix of corrected correlations as input for confirmatory factor analysis and structural equation modeling.

4.6. Confirmatory factor analysis

We assess the measurement properties of our latent constructs through Confirmatory Factor Analysis (CFA) in R 3.6.1 (R Core Team, 2019), using the lavaan package (Rosseel, 2012). Estimation of the measurement model was performed via maximum likelihood. While the model-implied covariance matrix differs from that observed in the sample ($X^2(224) = 337.692$, $p < .01$), the measurement model offers reasonable fit to the sample data as assessed by the Comparative Fit Index (CFI) value of 0.95, the Root Mean Square Error of Approximation (RMSEA) value of 0.06, the Standardized Root Mean Square Residual value of 0.06 and the Tucker-Lewis Index value of 0.94, all of which fall within common standards for acceptable fit (e.g., Kline, 2010) . We assess the psychometric properties of multi-item scales through an examination of CFA estimation results and comparison to established criteria for acceptability. Table 6 presents the item means and standard deviations, as well as scale reliabilities and factor loadings. The composite reliability (CR) of each scale exceeds the generally-accepted standard of 0.70 (Nunnally, 1978). Further, average variance extracted (AVE) for each construct exceeds the 0.50 threshold (Fornell & Larcker, 1981), providing evidence for convergent validity. Lastly, AVE for each latent construct was significantly greater than its squared correlation (Table 7), lending evidence for discriminant validity (Fornell & Larcker, 1981).

Table 6
Results of measurement model estimation.

Construct	Item	Standardized loading	*z*-value	P(>|z|)	Mean	SD
SPE risk mitigation effectiveness $\omega = 0.914$	SPE1	0.839	11.612	< 0.001	5.695	1.549
	SPE2	0.896	12.909	< 0.001	5.832	1.376
	SPE3	0.822	11.244	< 0.001	5.565	1.574
	SPE6	0.864	12.168	< 0.001	5.687	1.468
Rating Dissonance $\omega = 0.923$	RD1	0.907	13.053	< 0.001	2.504	1.935
	RD3	0.945	13.970	< 0.001	2.389	1.800
	RD4	0.757	10.001	< 0.001	1.740	1.120
Rating Justification $\omega = 0.892$	RJ2	0.903	12.863	< 0.001	5.855	1.222
	RJ3	0.823	11.152	< 0.001	5.649	1.341
	RJ5	0.850	11.702	< 0.001	5.656	1.329
Perceived Accuracy $\omega = 0.887$	A4	0.877	12.444	< 0.001	6.107	1.191
	A5	0.902	13.034	< 0.001	5.885	1.334
	A7	0.823	11.262	< 0.001	6.031	1.301
	A9	0.697	8.885	< 0.001	5.137	1.563
Fear of Supplier Dispute $\omega = 0.796$	FD1	0.686	8.293	< 0.001	2.481	1.931
	FD2	0.764	9.530	< 0.001	2.298	1.851
	FD4	0.834	10.706	< 0.001	1.878	1.489
Sufficiency of Requirement Definition $\omega = 0.926$	SRD1	0.954	14.605	< 0.001	5.496	1.470
	SRD2	0.973	15.116	< 0.001	5.496	1.366
	SRD4	0.793	10.851	< 0.001	4.924	1.639
Opportunism Attitude: Threat $\omega = 0.756$	L1	0.755	7.460	< 0.001	2.489	1.935
	L2	0.803	7.811	< 0.001	2.565	2.004
Opportunism Attitude: Debt $\omega = 0.729$	L3	0.776	8.180	< 0.001	1.557	1.117
	L4	0.744	7.894	< 0.001	1.740	1.298

5. Results

The structural model (Fig. 1) was estimated using structural equation modeling (SEM), via maximum likelihood. In addition to estimation of our hypothesized paths (as depicted within the figure), we estimated effects from a series of statistical controls on each endogenous variable to guard against bias due to potential confounds. First, a control variable (*group*), a dummy variable, accounts for variation in endogenous variables that may be attributed to consolidation of data across the two collections; responses collected through the latter survey were coded with a value of one. Second, we control for respondents' perceptions of purchase criticality (*criticality*) to account for any variation in *SPE risk mitigation effectiveness*, and mediating variables in the model, that might be attributed to the importance perceptions by respondents in the buying organization. To account for the nature of the relationship and the nature of the performance under evaluation, we include controls for contract duration (*duration*) and performance complexity (*complexity*). As with *criticality*, we control for their effects on mediating variables in our model as well as on *SPE risk mitigation effectiveness*. Lastly, we control for importance of the relationship to the firm by including a measure of contract value (*value*), again controlling all mediators as well as *SPE risk mitigation effectiveness*. We apply a natural-log transformation to *value*, which is measured in dollars, to improve normality (i.e., to reduce undue leverage from those in the tail of the distribution) and interpretability of coefficient estimates.

Fit indices for the structural model ($X^2(336) = 472.36$, $p < .01$) suggest adequate fit to the sample data (CFA = 0.94, RMSEA = 0.06, SRMR = 0.08, TLI = 0.93). Standardized estimates of path coefficients for hypothesized effects are presented within Table 8, along with their corresponding *z*-values. Non-significant effects ($p > .05$) were estimated from control variables *group, duration, complexity and value* on the endogenous values. Effects from control variable *criticality* were also non-significant on all endogenous variables, with the exceptions of *Fear of Supplier Dispute* ($B = -0.184$, $z = -2.208$, $p = .027$) and *SPE risk mitigation effectiveness* ($B = 0.159$, $z = 2.088$, $p = .037$). Returning to the hypothesized relationships, total effects on *SPE risk mitigation effectiveness* from mediated variables are presented within Table 9; to assess mediation we construct confidence intervals using the Monte Carlo method (Hayes & Scharkow, 2013; Preacher & Selig, 2012). The structural model explained 41.26% of the variation in *SPE risk mitigation effectiveness*, 51.79% of the variation in *rating justification*, 33.44% of the variation in *rating dissonance*, 51.99% of the variation in *perceived accuracy* and 47.80% of the variation in *fear of supplier dispute*. The model offered statistical support for each of the eight hypothesized relationships.

5.1. Effect of rating justification: a procedural rigor path

The path from sufficiency of requirement definition through perceived accuracy and rating justification to SPE risk mitigation effectiveness implicates the importance of procedural rigor and the contractual aspects of business, with a focus on thoroughly identifying requirements ex ante and on justified and pragmatic processing. Demonstrating stronger effect sizes than the path going through rating dissonance, this path indicates the supremacy of setting and following through on a shared, objective vision. Ex ante defining of requirements determines the perceived accuracy of assessment, and the more accurate the assessment, the stronger the outcome in terms of a justified rating and an efficacious supplier performance evaluation – one that can lend confidence of a future source selection team in its ability to assess the risk of adverse selection.

5.2. Effect of rating dissonance

Rating dissonance decreased SPE risk mitigation effectiveness, which was in turn increased by fear of supplier dispute. This path indicates greater fear of a supplier's dispute to the SPE associated with greater rating dissonance. Tracing the relationships back further, fear of supplier dispute is increased by the performance evaluator's opportunism attitude-debt.

Table 7
Correlations among latent constructs.

	Correlations							
	1.	2.	3.	4.	5.	6.	7.	8.
1. SPE risk mitigation effectiveness	**0.853**							
2. Rating dissonance	−0.302	**0.901**						
3. Rating justification	0.581	−0.184	**0.856**					
4. Perceived accuracy	0.525	−0.276	0.664	**0.815**				
5. Fear of supplier dispute	−0.369	0.556	−0.297	−0.448	**0.752**			
6. Sufficiency of requirement Definition	0.565	−0.291	0.507	0.653	−0.378	**0.899**		
7. Opportunism Attitude: Threat	−0.117	0.075	−0.110	−0.320	0.384	−0.165	**0.780**	
8. Opportunism Attitude: Debt	−0.133	0.386	−0.145	−0.270	0.561	−0.265	0.434	**0.758**

Square root of average variance extracted (AVE) shown in bold on the diagonal.

Table 8
Results of structural model estimation.

Direct effect	Standardized coefficient		z-value	P(>|z|)
H1	Rating Dissonance → SPE Risk Mitigation Effectiveness	−0.217	−2.641	0.008**
H2	Perceived Accuracy → Rating Justification	0.670	6.333	< 0.001**
H3	Rating Justification → SPE Risk Mitigation Effectiveness	0.544	5.342	< 0.001**
H4	Perceived Accuracy → Fear of Supplier Dispute	−0.288	−3.061	0.002**
H5	Fear of Supplier Dispute → Rating Dissonance	0.561	4.921	< 0.001**
H6	Sufficiency of Requirement Definition → Perceived Accuracy	0.635	6.747	< 0.001**
H7	Opportunism Attitude-Threat → Perceived Accuracy	−0.196	−2.413	0.016*
H8	Opportunism Attitude-Debt → Fear of Supplier Dispute	0.488	4.123	< 0.001**

* $p < .05$.
** $p < .01$.

Table 9
Total effects on SPE risk mitigation effectiveness from mediated variables.

Total effect	Standardized coefficient	95 percent confidence interval	
		Lower bound	Upper bound
Perceived Accuracy → SPE risk mitigation effectiveness	0.400	0.290	0.522
Fear of Supplier Dispute → SPE risk mitigation effectiveness	−0.121	−0.224	−0.033
Sufficiency of Requirement Definition → SPE risk mitigation effectiveness	0.255	0.172	0.353
Opportunism Attitude-Threat → SPE risk mitigation effectiveness	−0.079	−0.148	−0.016
Opportunism Attitude-Debt → SPE risk mitigation effectiveness	−0.059	−0.119	– 0.015

Note. Confidence intervals constructed using Monte Carlo simulation with 5000 draws.

5.3. Differential effects of two types of opportunism

The factor analysis confirmed a valid distinction between two forms of opportunism attitude – debt and threat – that are relevant in the context of SPE. The mean values of items measuring opportunism attitude-debt are noticeably lower than those measuring opportunism attitude-threat, indicating that the threat is a more established practice. Opportunism attitude-debt strongly increases the fear of a supplier dispute. Opportunism attitude-threat decreases the perceived accuracy of the SPE. Thus, the two forms of opportunism have differential effects on the paths to SPE risk mitigation effectiveness, and they ultimately weaken SPE risk mitigation effectiveness.

6. Discussion

The purpose of the research was to identify and explain the factors contributing to the risk-mitigating effectiveness of SPEs - the extent to which the supplier performance information collection and usage achieves the intended goal of mitigating the risk of adverse selection, and how SPE misuse undermines the effectiveness and efficiency of SPEs. A mixed method of qualitative interviews of buyers and suppliers followed by a survey of buy-side performance evaluators was employed in order to build and test a model of antecedents to SPE risk mitigation effectiveness with the following implications for research and practice.

6.1. Research implications

These empirical findings incriminate SPEs as vulnerable to unique manifestations of opportunism. SPE misuse undermines the powerful benefits of SPEs (see Table 10). Organizational behavior theory's explanatory power appears paramount over agency theory with regard to SPE risk mitigation effectiveness. This is extraordinary because organizational theory represents only 7% of theories relied upon in supply chain research (Defee et al., 2010). In a parallel to findings in organizational behavior literature regarding the impact on employee behavior of belief in the ratings portrayal (Ilgen et al., 1979), this research suggests that in order for SPEs to be effective in mitigating the risk of future adverse selection, consumers of the information (i.e., current suppliers and future buyers) must believe the SPE is true. This renders SPE accuracy and sufficient rating justifications into essential factors explaining SPE risk mitigation effectiveness. This finding is consistent with previous research that found performance data instability to impede supplier's reliance on the SPEs as a basis for improvements (Hald & Ellegaard, 2011). It also corresponds to a prior finding that buyers sometimes struggle to muster supporting information (i.e., the rating justification) to support a SPE (Hald & Ellegaard, 2011). Therefore, this research contributes to the stream of SPE research pertaining to functions of SPEs by quantitatively confirming prior qualitative findings and further by showing how – through accuracy and rating justification – SPEs can become more effective.

With regard to agency theory, two dimensions of agency operated simultaneously, and a third novel dimension emerged. First, the supplier is considered an agent of the buyer in promulgating the buyer's mission. Second, the buyer team is comprised of multiple agents to itself. In the case of multiple evaluators in different sub-organiza-

Table 10
Implications of SPE opportunism.

SPE benefit (Schmitz & Platts, 2003)	Opportunism implication
(1) Prioritize supplier improvement activities	Opportunistic behavior may divert activity to managing the performance evaluations rather than the actual contracted work activities. May result in reduced improvement activities, and discontinuities if performance evaluators change. Erosion of continuous improvement and misaligned buyer-seller priorities result.
(2) Focus management attention on critical suppliers	Opportunism diverts focus to non-operational and non-strategic priorities. A critical supplier that is mismanaged may become discouraged, not re-applying for future contract opportunities or diminishing investments such as top talent and technology. Inaccurate SPEs also mislead management attention, wasting resources and diminishing performance outcomes.
(3) Support supplier selection decisions	Unreliable or even questionable past performance data will be less likely to be used, representing not just a lost opportunity to improve supplier selection but also wasted resources. Where data corrupted by opportunism is used, it will lead to inaccurate supplier selection outcomes, with concomitant waste of resources and poorer quality outcomes.
(4) Communicate dissatisfaction with supplier performance	Lack of credible feedback mechanisms will erode supplier investment and motivation to improve performance. Perceived buyer susceptibility to opportunism may encourage suppliers to reciprocate or file a protest rather than accept negative feedback.
(5) Communicate performance expectations to suppliers	Use of threats, coercion, or imposing of a sense of debt will re-align supplier focus from performance expectations to managing expectations of individual performance evaluators.
(6) Document historical performance	SPE opportunism renders valueless any historical performance documentation. This represents a large waste of resources as well as a failure to comply by regulatory and ethical standards.
(7) Inform the purchasing department of supply base performance	Purchasing managers that rely on performance evaluation information corrupted by SPE opportunism will misallocate resources and arrive at misinformed supplier management strategies.
(8) Influence suppliers	Companies aggregate purchasing power to gain influence over suppliers through the power of incentives from bigger contracts as well as more stability and efficiency. To the extent that company priorities become disaggregated by opportunistic behavior at the individual level, company policies lose their influence. Buying company reputation suffers, which may discourage good suppliers from wanting to engage in future business with that buyer, or where already contracted, suppliers may not commit their best resources.
(9) Continuously improve	Inaccurate or untrustworthy feedback makes accurate identification of continuous improvement opportunities difficult. SPE opportunism shifts focus from process and service delivery improvement to "managing the performance evaluator".

tions, multiple agency relationships exist, and each can hold different interests. The third unsuspected dimension of agency pertains to the program (i.e., the purchased requirement such as a construction project or complex system). In some cases, both performance evaluators and supplier employees could begin to identify more with the program than with their employers. In other words, sometimes what is advantageous for the program can supersede what is advantageous for either the buyer team or the supplier. This explains the halo effect afforded a supplier who fails in one instance of performance, yet the evaluator does not mention the failure in the SPE because of reluctance to taint the program or the supplier's chance for future business.

This research finds two distinct types of attitudes toward opportunism in the context of SPE effectiveness, *threat* and *debt*. Buyers sometimes believe it is acceptable to engage in SPE misuse by weaponizing the SPE as leverage to reap concessions from suppliers. Consistent with prior research (Schmitz & Platts, 2004), buyers may do so by threatening a lower rating in exchange for a desired action. A novel finding is that buyers may also reap concessions by imposing a debt by "gifting" a more positive SPE than earned or documented. In either case, SPE misuse by the performance evaluator diminishes SPE accuracy and, paradoxically, harms the employer by compromising the organization's ability to avoid poor suppliers in the future. These findings, therefore, contribute to the SPE stream of literature pertaining to outcomes by illuminating a temporal dimension – SPE misuse by one employee serving as supplier performance evaluator (current agent) could inhibit the effectiveness of another employee serving on a source selection team (future agent).

Recent research finds that judgments of whether a particular act constitutes opportunism are subjective, dependent on the type of behavior, type of exchange, and personality (Arikan, 2018). Victims are more likely to perceive an act as opportunistic than are perpetrators. This may explain why some performance evaluators may hold an attitude that threatening a supplier is justifiable. It may also explain why some performance evaluators believe it acceptable to put suppliers in debt of repaying favors of positive SPEs. Performance evaluators simply do not perceive these behaviors as SPE misuse.

The findings from this research seem to confirm other aspects of SPE use from prior research. For example, Hald and Ellegaard (2011) found that buyers changed ratings based on reactions from suppliers implying that: (1) ratings were inaccurate and (2) buyers are concerned about supplier reactions. Similarly, this research found that SPE inaccuracy increases the fear of a supplier dispute to the ratings. The fear of a dispute creates dissonance among the buyer's evaluation team, which, in turn, decreases the risk mitigating effectiveness of SPEs. The implication is that performance evaluators often fail to correctly perceive SPE misuse even though SPE misuse propagates numerous negative consequences that undermine SPE effectiveness.

6.2. Practical implications

Rating justification had the strongest effect on SPE risk mitigation effectiveness. Thus, for those seeking to increase SPE risk mitigation effectiveness, efforts should be made to more thoroughly justify ratings—which implies an important effect from minimizing SPE misuse. This research offers insights as to how to improve rating justifications. First, buying organizations can address the effort required to justify a SPE. This can be done by making more time available to evaluators to conduct SPEs by hiring more evaluators, by dedicating evaluators to the task of SPE, or by reducing evaluators' non-SPE duties.

Some evidence suggests that buying organizations can improve rating justifications by sufficiently defining rating definitions. Doing so may require tailoring ratings and their definitions to the particular goods or services being procured rather than using common, generic definitions. Sourcing teams should further define performance criteria, how each will be measured, and develop thresholds for each that unambiguously lead to the specific performance ratings. These performance criteria and rating definitions should be defined in the request for proposals and requirements documents, and then set in the resulting contract in order for the supplier to more likely believe and accept the ratings as legitimate.

For the practicing manager, the model illustrates the importance of early steps in the process to create inputs to the rating system that are clearly defined. Buyers are notorious for omitting details and including ambiguous information. Greater perceived accuracy results when the buyer takes the time to clearly identify and communicate needs, and perceived accuracy, to an important extent, justifies supplier performance ratings.

Some performance evaluators believe that misuse of SPE ratings as leverage is acceptable—either (or both) as a threat to a supplier during performance and prior to a SPE or as a means to extract concessions post hoc from a supplier in exchange for a more favorable SPE rat-

ing. The survey data from performance evaluators was corroborated by testimonies from suppliers. Such an attitude toward opportunism was found to lower SPE accuracy. Since perceived accuracy was found to be a central construct leading to SPE risk mitigation effectiveness, SPE misuse as leverage should be explicitly addressed in hiring, training, and procurement policy.

Since the level of opportunism attitude-debt is lower than that of opportunism attitude-threat, there appears to be a greater acceptance among buyers of using the SPE as a threat for concessions from the supplier, perhaps because that, over time, the performance evaluator has learned that they can get what they want from suppliers. Further, the prevalence of a threat tactic versus a debt tactic suggests that buyers are more reluctant to use an overly generous SPE to create an obligation to return a favor.

6.3. Study limitations

Although similar to other studies of public procurement, the response rate is low calling into question the generalizability of the results. The response may have been subdued by several factors. First, past performance data is considered sensitive and proprietary information; some prospective respondents may have been uncomfortable participating. Furthermore, government agencies have discouraged the participation in surveys via cumbersome approvals. Another contributing factor may have been the data breach by the Office of Personnel Management involving the loss of sensitive information of 21 million government employees (Nelson & Tau, 2015). Additionally, survey solicitations via email originating outside of the government domain may be flagged as junk mail or spam. Finally, survey length was a likely culprit. Low response rate is not uncommon in business research. Melnyk, Page, Wu, and Burns (2012) revealed a steady decline of 1% annually since 2002. Five top journals reported low-end survey response rates ranging from 3% to 8%. Another limitation could include self-selection bias. Those respondents to the survey who were highly dissatisfied with the SPE policy or system could have been more inclined to respond to the survey. Absent information about non-respondents, non-response bias could only be inferred by comparing early to late respondents.

6.4. Future research directions

The trade-off of a current gain in supplier performance versus a future risk reduction in adverse selection raises an interesting research question. Researchers could explore what level of prospective gain will prompt an evaluator to make such a tradeoff and render the threat, and, conversely, what level of gain would be insufficient to warrant a threat? Future research could also explore the origins of opportunism attitudes that drive SPE misuse, particularly the debt form. Specifically, how much of the attitude that creating indebtedness to serve as leverage is an acceptable practice is held by the performance evaluator or by his or her supervisor (i.e., the reviewing official)? This answer would suggest to an ethically-minded organization the source in order to mitigate or eliminate the practice. Future research could also bolster our initial findings by examining actual acts of opportunism rather than attitudes.

Funding

This research was supported by a grant from the Naval Postgraduate School's Acquisition Research Program.

Appendix A. Measurement scale

Construct Scale Items [a]

SPE risk mitigation effectiveness

SPE1: This CPAR will help inform evaluators about this contractor's performance risk in a future source selection evaluation.

SPE2: If future source selection evaluators read this SPE, they can assess the risk of dealing with this contractor.

SPE3: This CPAR will reduce future source selection evaluators' uncertainty about this firm's likelihood of performing similar work well.

SPE6: With this SPE, future source selection evaluators can be confident in their assessment of the risk of this contractor successfully performing on a similar future contract.

Rating dissonance

RD1: Between myself, the "Reviewing Official," and other performance evaluators, there was some disagreement on at least one SPE rating.

RD2: Significant effort was required to deliberate with others as to what rating(s) to assign.

RD4: The government team had difficulty reaching consensus on the ratings or narrative justification.

Rating justification

RJ2: The rationale for the assigned SPE rating was thoroughly documented.

RJ3: An inspector general would conclude that the SPE rating was sufficiently explained.

RJ5: Anyone who reads this SPE will understand the ratings based on the supporting information in the report.

Perceived accuracy

A4: The government accurately measured the contractor's performance level.

A5: The government consistently measured the contractor's performance level.

A7: All of the assessed ratings could be traced back to records of contractor performance.

A9: Had ten other qualified people completed this SPE, each would have arrived at the exact same ratings.

Fear of supplier dispute

FD1: If I reported the contractor's performance accurately, the contractor would have disputed/rebutted the rating(s).

FD2: While completing the SPE, at least one member of the government team was concerned that the contractor might dispute the assigned rating.

FD4: To report the ratings that the contractor actually deserved would have consumed too much time responding to the contractor's rebuttal.

Sufficiency of requirement definition

SRD1: The requirement was very well defined in the contract/task order/delivery.

SRD2: The contract/task order/delivery order (including the statement of work, performance work statement, specification, drawings, etc.) defined the requirement very well.

SRD4: There were no ambiguities in the definition of the requirement (including the statement of work, performance work statement, specification, drawings, etc.).

Opportunism attitude-threat

OT1: It is ok for the Government to threaten the contractor with a lower SPE rating.

OT2: It is ok for the Government to use the SPE as bargaining leverage with the contractor.

Opportunism attitude-debt

OD3: If we give the contractor a SPE that is better than what they deserve, the contractor should reciprocate in some way.

OD4: Leverage can be gained by providing the contractor an overly favorable SPE.

[a] All responses were obtained using 7-point Likert-type scales.

References

Aberdeen Group (2005). The supplier performance measurement benchmark report. Boston, Massachusetts, September.

Anderson, E (1988). Transaction costs as determinants of opportunism in integrated and independent sales forces. Journal of Economic Behavior & Organization, 9(3), 247–264.

Arikan, A T (2018). Opportunism is in the eye of the beholder: Antecedents of subjective opportunism judgments. Journal of Business Ethics, 1–17. doi:10.1007/s10551-018-3873-7.

Armstrong, J S, & Overton, T S (1977). Estimating nonresponse bias in mail surveys. Journal of Marketing Research, 14(3), 396–402.

Ashnai, B, Henneberg, S C, Naudé, P, & Francescucci, A (2016). Inter-personal and inter-organizational trust in business relationships: An attitude–behavior–outcome model. Industrial Marketing Management, 52, 128–139.
Azadegan, A (2011). Benefiting from supplier operational innovativeness: The influence of supplier evaluations and absorptive capacity. Journal of Supply Chain Management, 47(2), 49–64.
Barney, J B, & Hesterly, W (2006). Organizational economics: Understanding the relationship between organizations and economic analysis. The SAGE Handbook of Organization Studies, 111–148.
Barney, J B, & Ouchi, W G (1986). Organizational Economics. Jossey-Bass.
Beausoleil, J W (2010). Past Performance Handbook: Applying Commercial Practices to Federal Procurement (2 ed.). Vienna, VA: Management Concepts.
Berrios, R (2006). Government contracts and contractor behavior. Journal of Business Ethics, 63, 119–130.
Bettman, J R (1973). Perceived risk and its components: A model and empirical test. Journal of Marketing Research, 10(May), 184–190.
Brown, A, Inceoglu, I, & Lin, Y (2017). Preventing rater biases in 360-degree feedback by forcing choice. Organizational Research Methods, 20(1), 121–148.
Buffa, F P, & Ross, A D (2011). Measuring the consequences of using diverse supplier evaluation teams: A performance frontier perspective. Journal of Business Logistics, 32(1), 55–68.
Campbell, D J, & Lee, C (1988). Self-appraisal in performance evaluations: Development versus evaluation. Academy of Management Review, 13(2), 302–314.
CAPS Research (2011). Supplier quality & delivery performance; 2011 supply management benchmarking report. Tempe, AZ: Institute for Supply Management and W.P. Carey School of Business at Arizona State University.
Cavinato, J, & Kauffman, R G (2000). The purchasing handbook: A guide for the purchasing and supply professional. New York: McGraw Hill.
Churchill, G A, Jr. (1979). A paradigm for developing better measures of marketing constructs. Journal of Marketing Research, 16(1), 64–73.
Combs, J G, & Ketchen, D J, Jr. (1999). Explaining interfirm cooperation and performance: Toward a reconciliation of predictions from the resource-based view and organizational economics. Strategic Management Journal, 20(9), 867–888.
Cormican, K, & Cunningham, M (2007). Supplier performance evaluation: Lessons from a large multinational organization. Journal of Manufacturing Technology Management, 18(4), 352–366.
Cousins, P D, Lawson, B, & Squire, B (2008). Performance measurement in strategic buyer-supplier relationships: The mediating role of socialization mechanisms. International Journal of Operations and Production Management, 28(3), 238–258.
D'Arcy, J, & Devaraj, S (2012). Employee misuse of information technology resources: Testing a contemporary deterrence model. Decision Sciences, 43(6), 1091–1124.
Day, M, Fawcett, S E, Fawcett, A M, & Magnan, G M (2013). Trust and relational embeddedness: Exploring a paradox of trust pattern development in key supplier relationships. Industrial Marketing Management, 42(2), 152–165.
Deb, J, Li, J, & Mukherjee, A (2016). Relational contracts with subjective peer evaluations. The Rand Journal of Economics, 47(1), 3–28.
Defee, C, Williams, B, Randall, W S, & Thomas, R (2010). An inventory of theory in logistics and SCM research. International Journal of Logistics Management, 21(3), 404–489.
Dillman, D (2000). Mail and Internet Surveys: The Tailored Design Method (2nd ed.). New York: John Wiley and Sons, Inc..
Disdier, A-C, & Head, K (2008). The puzzling persistence of the distance effect on bilateral trade. The Review of Economics and Statistics, 90(1), 37–48.
Donaldson, L (1990). The ethereal hand: Organizational economics and management theory. Academy of Management Review, 15(3), 369–381.
Dowst, S (1972). How designers rate suppliers—And purchasing. Purchasing, 12(June), 87–93.
Eckerd, S, Hill, J, Boyer, K K, Donohue, K, & Ward, P T (2013). The relative impact of attribute, severity, and timing of psychological contract breach on behavioral and attitudinal outcomes. Journal of Operations Management, 31(7–8), 567–578.
Eisenhardt, K (1989). Agency theory: An assessment and review. Academy of Management Review, 14(1), 57–74.
Erez, A, Schilpzand, P, Leavitt, K, Woolum, A H, & Judge, T A (2015). Inherently relational: Interactions between peers' and individuals' personalities impact reward giving and appraisal of individual performance. Academy of Management Journal, 58(6), 1761–1784.
Flint, D J, Woodruff, R B, & Gardial, S F (2002). Exploring the phenomenon of customers' desired value change in a business-to-business context. Journal of Marketing, 66(4), 102–117.
Folger, R, Konovsky, M A, & Cropanzano, R (1992). A due process metaphor for performance appraisal. In Stawand, B M, & Cummings, L L (Eds.), Research in Organizational Behavior (Vol. 13, pp. 129–177). Greenwich, CT: JAI Press.
Fornell, C, & Larcker, D (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50.
Frohlich, M T, & Westbrook, R (2001). Arcs of integration: An international study of supply chain strategies. Journal of Operations Management, 9(2), 185–200.
Giannakis, M (2007). Performance measurement of supplier relationships. Supply Chain Management: An International Journal, 12(6), 400–411.
Gibbons, R, & Roberts, J (Eds.). (2013). The handbook of organizational economics. Princeton University Press.
Gimenez, C, & Sierra, V (2013). Sustainable supply chains: Governance mechanisms to greening suppliers. Journal of Business Ethics, 116(1), 189–203.
Gonzalez-Padron, T, Hult, G T M, & Calantone, R (2008). Exploiting innovative opportunities in global purchasing: An assessment of ethical climate and relationship performance. Industrial Marketing Management, 37(1), 69–82.
Gordon, S (2008). Supplier evaluation and performance excellence. Fort Lauderdale, Florida: J. Ross Publishing.
Gouldner, A (1960). The norm of reciprocity: A preliminary statement. American Sociological Review, 25, 161–178.
Government Accountability Office (GAO) (2002). Contract management: Guidance needed for using performance-based service contracting (GAO-02-1049). Washington, DC.
Government Accountability Office (GAO) (2007). Defense acquisitions: Improved management and oversight needed to better control DoD's acquisition of services (GAO-07–832T) Washington, DC.
Government Accountability Office (GAO) (2009). Federal contractors: better performance information needed to support agency contract award decisions (GAO-09-1032). Washington, DC: Government Accountability Office.
Government Accountability Office (GAO) (2014). Contractor performance: actions taken to improve reporting of past performance information (GAO-14-707). Washington, DC: Government Accountability Office.
Granovetter, M S (1985). Economic action and social structure: The problem of embeddedness. American Journal of Sociology, 9(3), 481–510.
Greenberg, M S (1980). A theory of indebtedness. Social Exchange (pp. 3–26). Boston, MA: Springer.
Hair, J, Black, W, & Babin, B (2010). Multivariate Data Analysis (7th ed.). Upper Saddle River, New Jersey: Prentice Hall.
Hald, K S, & Ellegaard, C (2011). Supplier evaluation processes: The shaping and reshaping of supplier performance. International Journal of Operations and Production Management, 31(8), 888–910.
Handfield, R B, & Nichols, E L, Jr. (2004). Key issues in global supply base management. Industrial Marketing Management, 33(1), 29–35.
Handfield, R B, Cousins, P D, Lawson, B, & Petersen, K J (2015). How can supply management really improve performance? A knowledge-based model of alignment capabilities. Journal of Supply Chain Management, 51(3), 3–17.
Hawkins, T, Berkowitz, D, Muir, W, & Gravier, M (2015). Improving services supply management in the defense sector: How the procurement process affects B2B service quality. Journal of Purchasing and Supply Management, 21, 81–94.
Hawkins, T, & Muir, W (2014). An exploration of knowledge-based factors affecting procurement compliance. Journal of Public Procurement, 14(1), 1–32.
Hawkins, T G, Wittmann, C M, & Beyerlein, M M (2008). Antecedents and consequences of opportunism in buyer–supplier relations: Research synthesis and new frontiers. Industrial Marketing Management, 37(8), 895–909.
Hayes, A F, & Scharkow, M (2013). The relative trustworthiness of tests of indirect effects in statistical mediation analysis. Does method really matter? Psychological Science, 24, 1918–1927.
van der Heijden, B, & Nijhof, A (2004). The value of subjectivity: Problems and prospects for 360-degree appraisal systems. International Journal of Human Resource Management, 15(3), 493–511.
Hill, J A, Eckerd, S, Wilson, D, & Greer, B (2009). The effect of unethical behavior on trust in a buyer–supplier relationship: The mediating role of psychological contract violation. Journal of Operations Management, 27, 281–293.
Hu, H Y, Chiu, S I, Yen, T-M, & Cheng, C-C (2015). Assessment of supplier quality performance of computer manufacturing industry by using ANP and DEMATEL. The TQM Journal, 27(1), 122–134.
Huo, B (2012). The impact of supply chain integration on company performance: An organizational capability perspective. Supply Chain Management: An International Journal, 17(6), 596–610.
Husser, J, Gautier, L, André, J M, & Lespinet-Najib, V (2014). Linking purchasing to ethical decision-making: An empirical investigation. Journal of Business Ethics, 123(2), 327–338.
Ilgen, D R, Fisher, C D, & Taylor, M S (1979). Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, 64(4), 349–371.
Jehn, K A, & Chatman, J A (2000). The influence of proportional and perceptual Conflictcomposition on Team performance. International Journal of Conflict Management, 11(1), 56.
John, G (1984). An empirical investigation of some antecedents of opportunism in a marketing channel. Journal of Marketing Research, 21(3), 278–289.
Kaufmann, L, & Wagner, C M (2017). Affective diversity and emotional intelligence in cross-functional sourcing teams. Journal of Purchasing and Supply Management, 23(1), 5–16.
Kelman, S (2010). Three ways to revitalize the use of past-performance data. Federal Computer Week. 13 May, available at http://fcw.com/blogs/lectern/2010/05/steve-kelman-past-performance-source-selection.aspx (accessed on 15 August 2017).
Kingshott, R (2006). The impact of psychological contracts upon trust and commitment within supplier–buyer relationships: A social exchange view. Industrial Marketing Management, 35(6), 724–739.
Kinicki, A J, Prussia, G E, McKee-Ryan, F M, & Wu, B (2004). A covariance structure analysis of employees' response to performance feedback. Journal of Applied Psychology, 89(6), 1057–1069.
Kline, R B (2010). Principles and practice of structural equation modeling. New York: Guilford Press.
Korschun, D (2015). Boundary-spanning employees and relationships with external stakeholders: A social identity approach. Academy of Management Review, 40(4), 611–629.
Kuo, H, & Nakhata, C (2016). Price promotions and products with low consumer ratings. The Journal of Consumer Marketing, 33(7), 517–527.
Lages, L, Silva, G, & Styles, C (2009). Relationship capabilities, quality, and innovation as determinants of export performance. Journal of International Marketing, 17(4), 47–70.

Lam, P K, Chin, K S, & Pun, K F (2007). Managing conflict in collaborative new product development: A supplier perspective. The International Journal of Quality & Reliability Management, 24(9), 891–907.
Leuschner, R, Rogers, D S, & Charvet, F F (2013). A meta-analysis of supply chain integration and firm performance. Journal of Supply Chain Management, 49(2), 34–57.
Levy, P, Cawley, B, & Foti, R (1998). Reactions to appraisal discrepancies: Performance ratings and attributions. Journal of Business and Psychology, 12(4), 437–455.
Lorentz, H, Töyli, J, Solakivi, T, Hälinen, H M, & Ojala, L (2012). Effects of geographic dispersion on intra-firm supply chain performance. Supply Chain Management: An International Journal, 17(6), 611–626.
Lovelace, K, Shapiro, D L, & Weingart, L R (2001). Maximizing cross-functional new product teams' innovativeness and constraint adherence: A conflict communications perspective. Academy of Management Journal, 44(4), 779–793.
Luthans, F, & Youssef, C M (2007). Emerging positive organizational behavior. Journal of Management, 33(3), 321–349.
Maestrini, V, Maccarrone, P, Caniato, F, & Luzzini, D (2018). Supplier performance measurement systems: Communication and reaction modes. Industrial Marketing Management, 74, 298–308.
Malhotra, N K, Kim, S S, & Patil, A (2006). Common method variance in IS Research: A comparison of alternative approaches and a reanalysis of past Research. Management Science, 52(12), 1865–1883.
Melnyk, S, Page, T, Wu, S, & Burns, L A (2012). Would you mind completing this survey: Assessing the state of survey research in supply chain management. Journal of Purchasing and Supply Management, 18(1), 35–45.
Molm, L D (1997). Risk and power use: Constraints on the use of coercion in exchange. American Sociological Review, 62(1), 113–133.
Monczka, R M, Choi, T Y, Kim, Y, & McDowell, C (2011). Supplier relationship management: An implementation framework. Tempe, AZ: CAPS Research.
Monczka, R M, Trent, R J, & Handfield, R B (2002). Purchasing and supply chain management (2nd ed.). Cincinnati, Ohio, United States: South-Western.
Morgan, R, & Hunt, S (1994). The commitment-trust theory of relationshp marketing. Journal of Marketing, 58(3), 20–38.
Mount, M, Judge, T, Scullen, S, Sytsma, M, & Hezlett, S (1998). Trait, rater and level effects in 360-degree performance ratings. Personnel Psychology, 51(3), 557–576.
Nair, A, Jayaram, J, & Das, A (2015). Strategic purchasing participation, supplier selection, supplier evaluation and purchasing performance. International Journal of Production Research, 53(20), 6263–6278.
Nelson, C, & Tau, B (2015). OPM director resigns after massive personnel data hack. Dow Jones News. 10 July, available at http://www.advfn.com/news_OPM-Director-Resigns-After-Massive-Personnel-Data_67659874.html (accessed on 22 April 2019).
Nunnally, J (1978). Psychometric theory. New York: McGraw-Hill.
Office of Federal Procurement Policy (OFPP) (2000, May). Best practices for collecting and using current and past performance information. Washington, DC: Office of Management and Budget.
Office of Federal Procurement Policy (OFPP) (2011). Improving contractor past performance assessments: Summary of the Office of Federal Procurement Policy's review, and strategies for improvement [memorandum], January 21. Washington, DC: Office of Management and Budget.
Office of Personnel Management (2014). Common characteristics of the government fiscal year 2014. Washington, DC: Office of Personnel Management.
Palmatier, R (2008). Interfirm relational drivers of customer value. Journal of Marketing, 72(4), 76–89.
Paulhus, D (1991). Measurement and control of response Bias. In Robinson, J P, Shaver, P R, & Wrightsman, L S (Eds.), Measures of Personality and Social Psychological Attitudes: Volume I of Measures of Social Psychological Attitudes (2, pp. 17–59). San Diego: Academic Press, Inc..
Pelser, J, de Ruyter, K, Wetzels, M, Grewal, D, Cox, D, & van Beuningen, J (2015). B2B channel partner programs: Disentangling indebtedness from gratitude. Journal of Retailing, 91(4), 660–678.
Prahinski, C, & Benton, W (2004). Supplier evaluations: communication strategies to improve supplier performance. Journal of Operations Management, 22(1), 39–62.
Prahinski, C, & Fan, Y (2007). Supplier evaluations: The role of communication quality. Journal of Supply Chain Management, 43(3), 16–28.
Preacher, K J, & Selig, J P (2012). Advantages of Monte Carlo confidence intervals for indirect effects. Communication Methods and Measures, 6, 77–98.
Prior, D D (2012). The effects of buyer–supplier relationships on buyer competitiveness. Journal of Business & Industrial Marketing, 27(2), 100–114.
Prior, D D (2016). The impact of service worker personal resources on relationship quality in business solutions. Industrial Marketing Management, 53, 216–225.
R Core Team (2019). R: A langauge and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
Randall, D M, & Fernandez, M F (1991). The social desirability response Bias in ethics Research. Journal of Business Ethics, 10(11), 805–817.
Reilly, T, Saini, A, & Skiba, J (2018). Ethical purchasing dissonance: Antecedents and coping behaviors. Journal of Business Ethics, 1–21.
Rindfleisch, A, & Heide, J B (1997). Transaction cost analysis: past, present, and future applications. Journal of Marketing, 61(4), 30–54.
Rosseel, Y (2012). Lavaan: An R package for structural equation modeling. Journal of Statistical Software, 48(2), 1–36.
Saastamoinen, J, Reijonen, H, & Tammi, T (2017). The role of training in dismantling barriers to sme participation in public procurement. Journal of Public Procurement, 17(1), 1–30.
Saini, A (2010). Purchasing ethics and inter-organizational buyer–supplier relational determinants: A conceptual framework. Journal of Business Ethics, 95(3), 439–455.
Samaha, S A, Palmatier, R W, & Dant, R P (2011). Poisoning relationships: Perceived unfairness in channels of distribution. Journal of Marketing, 75(3), 99–117.
Schleper, M C, Blome, C, & Wuttke, D A (2017). The dark side of buyer power: Supplier exploitation and the role of ethical climates. Journal of Business Ethics, 140(1), 97–114.
Schmitz, J, & Platts, K W (2003). Roles of supplier performance measurement: Indication from a study in the automotive industry. Management Decision, 41(8), 711–721.
Schmitz, J, & Platts, K W (2004). Supplier logistics performance measurement: Indications from a study in the automotivs industry. International Journal of Production Economics, 89(2), 231–243.
Schoenherr, T, Modi, S B, Benton, W C, Carter, C R, Choi, T Y, Larson, P D, … Wagner, S M (2012). Research opportunities in purchasing and supply management. International Journal of Production Research, 50(16), 4556–4579.
Shapiro, S (2005). Agency theory. Annual Review of Sociology, 31, 263–284.
Shrauger, J, & Osberg, T (1981). The relative accuracy of self-predictions and judgments by others in psychological assessment. Psychological Bulletin, 90(2), 322–351.
Simpson, P M, Siguaw, J A, & White, S C (2002). Measuring the performance of suppliers: An analysis of evaluation processes. Journal of Supply Chain Management, 38(1), 29–41.
Steinle, C, Schiele, H, & Ernst, T (2014). Information asymmetries as antecedents of opportunism in buyer-supplier relationships: Testing principal-agent theory. Journal of Business-to-Business Marketing, 21(2), 123–140.
Tammi, Timo, Saastamoinen, Jani, & Reijonin, Helen (2014). Market orientation and SMEs' activity in public sector procurement participation. Journal of Public Procurement, 14(3), 304–327.
Tangpong, C, Li, J, & Hung, K T (2016). Dark side of reciprocity norm: Ethical compromise in business exchanges. Industrial Marketing Management, 55, 83–96.
Thomas, S, & Bretz, R (1994). Research and practice in performance appraisal: Evaluating employee performance in America's largest companies. SAM Advanced Management Journal, 59(2), 28–34.
Trombetta, W L, & Wilson, T L (1975). Foreseeability of misuse and abnormal use of products by the consumer. Journal of Marketing, 39(3), 48–55.
Ulaga, W (2003). Capturing value creation in business relationships: A customer perspective. Industrial Marketing Management, 32(8), 677–693.
Van der Valk, W, & Rozemeijer, F (2009). Buying business services: Towards a structured service purchasing process. Journal of Services Marketing, 23(1), 3–10.
P. Van Ecke G. Skouma W. Freund M. Goeskjaer B. Ooms Legal study on unfair commercial practices within b2b e-markets: Final report (European Commission Study ENTR/04/69)Retrieved fromhttp://ec.europa.eu/growth/content/legal-study-unfair-commercial-practices-b2b-e-markets-0_nn2006on 1 July 2019
Vargo, S L, & Lusch, R F (2011). It's all B2B… and beyond: Toward a systems perspective of the market. Industrial Marketing Management, 40(2), 181–187.
Wang, X, & Yang, Z (2013). Inter-firm opportunism: A meta-analytic review and assessment of its antecedents and effect on performance. Journal of Business & Industrial Marketing, 28(2), 137–146.
Wathne, K H, & Heide, J B (2000). Opportunism in interfirm relationships: Forms, outcomes, and solutions. Journal of Marketing, 64(4), 36–51.
Wieters, D, & Ostrom, L (1979). Supplier evaluation as a new marketing tool. Industrial Marketing Management, 8(2), 161–166.
Williamson, O E (1975). Markets and hierarchies: Analysis and antitrust implications. New York, NY: Free Press.
Williamson, O E (1981). The economics of organization: The transaction cost approach. American Journal of Sociology, 87(3), 548–577.
Williamson, O E (1987). Transaction cost economics: The comparative contracting perspective. Journal of Economic Behavior & Organization, 8(4), 617–625.
Williamson, O E (1993). Calculativeness, trust, and economic organization. Journal of Law and Economics, 36(1, Part 2), 453–486.
Worsham, J, Eisner, M, & Ringquist, E (1997). Assessing the assumptions: A critical analysis of agency theory. Administration and Society, 28(4), 419–440.
Yin, R K (2009). Case study research: Design and methods. Los Angeles, CA: Sage Publications.
Young, A L, & Quan-Haase, A (2013). Privacy protection strategies on Facebook: The internet privacy paradox revisited, *Information*. Communications Society, 16(4), 479–500.
Zeydan, M, Çolpan, C, & Çobanoğlu, C (2011). A combined methodology for supplier selection and performance evaluation. Expert Systems with Applications, 38(3), 2741–2751.